Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Is Stock BBS Content Correlated with the Stock Market?—A Japanese Case

Ken Maruyama
Next Solutions Inc.
Tokyo, Japan

Eiichi Umehara
Nomura Research Institute Inc
Tokyo, Japan

Hirohiko Suwa
The University of Electro-Communications
The Graduate School of Information Systems
Tokyo, Japan

Toshizumi Ohta
The University of Electro-Communications
The Graduate School of Information Systems
Tokyo, Japan

Abstract— **We analyze the relations between the stock market and a stock bulletin board system (BBS) in Japan. Previous studies in the USA found that the characteristics of messages posted on stock BBSs can predict market volatility and trading volume. We develop hypotheses based on the results of those analyses and apply statistical analysis to the data about companies mentioned in a large number of messages posted on the Yahoo! stock message board in Japan in 2005–2006. We analyze the contents of these messages using natural language processing. We find a significant correlation between the number of postings and market volatility and trading volume, and also find significant correlation between the amount of bullish and bearish opinion and the stock return.**

Keywords—**stock market, stock BBS, support vector machine, natural language processing, machine learning**

I. INTRODUCTION

The amount of consumer generated media posted by users on the Internet in Japan has increased in recent years. For instance, ASCII [8] reported the relation between weblogs and stocks. Moreover, Nikkei news [Dec 30, 2006] reported that: "When the reason for stock price falls was sought, it was found to be a strange message posted on a stock BBS. " Van Bommel [6] proposed a model for stock price overshoots caused by rumor; his model was based on the individual information diffusion model. He concluded that stock prices can change because of rumors on the Internet.

We analyze the relation between messages posted on an Internet stock message board and the stock market in Japan. We analyze the message content by natural-language processing. From the results, we derive the relation between bulletin board system (BBS) indexes (the number of messages and the message content) and the stock market (stock return, trading volume, and market volatility). We analyze in this paper the Japanese Yahoo! stock message board[1] (hereinafter called Yahoo! BBS).

II. PREVIOUS RESEARCH

In the USA, Antweiler and Frank [1] analyzed the number and content of more than 1,500,000 messages posted on Yahoo! and Raging Bull[2] concerning 45 companies on the Dow Jones industrial stock index and the Dow Jones Internet index by using natural-language processing (Naive Bayes) and empirically examined the relation between message boards and the stocks market. They found that (i) the message boards do not estimate the stock return, (ii) because the difference between bullish and bearish opinions drives stock trading, the message boards forecast the trading volume, and (iii) the message boards forecast the volatility of the next day's trading.

Harris and Raviv [4] proposed a mathematical model of speculative trading and analyzed the relation between the differences in opinion among traders and the trading volume and market volatility (absolute stock price change). In this model, speculative trading is caused by disagreements in opinion resulting from different interpretations of publicly available information. They pointed out that the absolute value of the stock price change and the trading volume has a positive correlation and that there is also a positive correlation between the absolute value of the forecast change in the final profit and loss by investor and volume.

Wyscocki [7] examined the cross-sectional and time-series determinants of the volume of messages posted on stock message boards on the Web. Using a sample of over 3000 stocks listed on Yahoo! message boards, he found that the cumulative posting volume is highest for firms with high short-seller activity, high market valuations relative to fundamentals, low institutional holdings, high trading volume, extreme performance, and high analyst following. Changes in daily posting volume are associated with earnings-announcements and daily changes in stock trading volume and returns. The overnight message-posting volume is found to

[1] Yahoo! stock message board is linked with Yahoo! Finance and has one topic for one stock. http://quote.yahoo.co.jp/

[2] Raging Bull is an Internet message board in United States where topics concerning stocks, business, and politics are discussed. http://ragingbull.quote.com/

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

predict changes in the next day's stock trading volume and returns, but it is difficult to obtain economical profit when commission is taken into consideration.

Tumarkin and Whitelaw [5] examined the causal relation between the number of and opinions in messages posted on Raging Bull and the return and trading volume by using an event study and a multi-auto-regression analysis. They concluded that the messages cannot estimate stock return and volume, and this result shows that the market works efficiently.

Yamashita et al. [11] analyzed the characteristics of messages posted on Yahoo! BBS in Japan. They pointed out that the hot news topics strongly influence the number of messages posted on this BBS. However, they analyzed only the relation between the number of messages posted on the BBS and the changes in stock prices and did not analyze the message contents.

III. TREND OF STOCK BBS IN JAPAN

Here, we describe the trend of Yahoo! BBS in Japan. We chose to analyze this BBS because message posting on it is very active in Japan[3]. The average number per company is 4034, i.e., 5.5 messages per day from Jan 1, 2005 to Dec 31, 2006 concerning 1541 companies listed on the Tokyo Stock Exchange 1st section[4]. However, some companies have no messages about them posted on the BBS, while others have many. Therefore, we analyze the relation between the BBS and the stock market for the top 50 companies in terms of the number of posted messages.

In 2005, the Nikkei 225 rose from 11,517.75 on January 4 to 16,111.43 on December 30 (39.9%). In 2006, it rose from 16,361.54 on January 4 to 17,225.83 on December 29 (5.28%). We guess that most of the 2005 messages were posted during a bullish market, whereas 2006 messages included some posted during a box market (a stock price is moving up and down repeating between an upper bound and a lower bound). However, we think that our analysis data include few messages posted during a bearish market.

On Yahoo! BBS, each message has a timestamp. Looking at the temporal distribution of the number of messages on the basis of timestamps reveals that there are many messages from 0900 to 1500 (while the market is open) and also from 1500 to 2400 (after the market has closed), but few overnight from 0000 to 0900 the next morning (before the market opens). Though details are omitted because of space limitations, there is a clear tendency for many postings to occur while the market is open and after it has closed and for very few to occur before it reopen. This is also true for other companies (data not shown). Therefore, we analyze messages by dividing them into three time periods (windows) (0000–0900, 0900–1500, and 1500–2400).

[3]A few BBSs have a message board for each company which can be watched and posted to by many unspecified people in Japan. New BBSs are now available (e.g., stock for everyone (established in April 2007: http://minkabu.jp/), but Yahoo! BBS has been operating since July 1998.

[4]We collect messages on BBSs using our original software. We use Toyokeizai Stock Price CD-ROM 2006 as market data.

IV. HYPOTHESIS SETTING

A lot of factors influence stock prices, such as fundamentals, analyst reports, and mass media reports. However, we focus on only an analysis of the relations between messages on a BBS and the stock market.

A. Relations between the number of posted messages and the stock market

[7] reported that the number of posted messages can estimate the next day's excess return. However, [1] concluded that the number of postings can not estimate the stock return. For the relation between the number of messages and trading volume, [7] showed that the number of overnight messages can explain the trading volume on the next day. For the relation between the number of messages and volatility, [1] showed that the number of messages has a positive correlation with volatility. Based on these previous research findings, we set up the following three hypotheses about the relation between the number of messages and the stock market.

Hypothesis 1: The number of messages has no correlation with the stock return.

The content of posted messages suggests that people who post on Internet BBSs may not be institutional investors but individuals interested in stocks. There are various sentiments such as "buy", "strong buy", "sell", "strong sell", and "hold". Therefore, the number of messages may be neutral with respect to bullish (stock price rise) and bearish (stock price fall). Thus, the number of messages might not have a correlation with the stock return.

Hypothesis 2: The number of messages correlates with trading volume.

If the people posting on the Internet BBS are investors who use online trading, they might may make investment judgments based on the posted messages and buy or sell those stocks. If the BBS posting becomes active and the number of messages increases, the sentiments of many investors might be reflected on the BBS, and the BBS could influence the investment behavior of individual investors and the trading volume might increase. That is, the number of messages might be a variable representing the degree to which individual investors using the Internet are interested in the stocks.

Hypothesis 3: The number of messages correlates with volatility.

As described in hypothesis 2, a BBS might influence investment behavior. As a result, the trading volume and the volatility of stock prices might increase.

B. Relation between contents of messages and stock market

As BBS indexes of the contents of messages, we focus on two indexes: the level of one-sided opinion of a bullish or bearish nature (hereinafter called the bullishness) and the extent to which opinions agree or disagree (hereinafter called the agreement index) during a certain period. Bullishness is the difference between the numbers of bullish and bearish messages during a certain period.

[1] claimed that there is a statistically significant correlation

between bullishness and stock return during the same time period. They showed that the relation between bullishness and trading volume and volatility is significant. [4] developed a model of trading in speculative markets based on differences among traders. They showed that when opinions disagree, absolute price changes and volume may increase. [1] showed that the agreement index has a correlation with volume. Because we could not find any previous research supporting this, we decide to test the hypothesis that there is no correlation with the agreement index and the stock return.

Hypothesis 4: Bullishness is positively correlated with stock return.

Hypothesis 5: Bullishness is positively correlated with volume.

Hypothesis 6: Bullishness is positively correlated with volatility.

If a BBS were to influence the investment behavior of individual investors, they might buy stock if there were a lot of bullish opinions on the BBS and sell if there were many bearish opinions. Thus, we want to investigate whether bullishness has a positive influence on stock return, volume, and volatility.

Hypothesis 7: The agreement index is not positively correlated with stock return.

Even if BBS message contents agree with the bullish or bearish market, the extent to which opinions agree or disagree might not be related to stock return.

Hypothesis 8: The agreement index is negatively correlated with volume.

Hypothesis 9: The agreement index is negatively correlated with volatility.

The model of [4] suggests that the differences in opinion among traders concerning the interpretation of public information lead to an increase in trading volume and volatility (absolute value of stock price change).

V. ANALYSIS METHOD

To examine these hypotheses, we analyze the Yahoo! BBS in Japan from Jan 1, 2005 to Dec 31 2006 (hereinafter the analysis period) for the top 50 companies in terms of the number of messages posted on Yahoo! BBS and listed on the Tokyo Stock Exchange 1st section.

A. Stock Markets

We analyzed five indexes: stock price return, excess price return, volume, volatility, and absolute value of return. The daily price return: R(t) is closing price(t)/closing price (t-1)-1. R(t) minus the return of Tokyo Stock Price Index (TOPIX) is the excess price return. We define the volatility as the standard deviation of returns for five days such as today plus and minus two days.

B. Collection of messages on Yahoo! BBS and classification into bullish and beaish messages.

We classify messages on Yahoo! BBS into three types: "bullish", "neither", and "bearish" using natural-language processing. First, a feature vector is extracted from each message using a morphological analyzer. To classify these messages, we analyze these vectors using support vector regression (SVR) [2]. Our procedure is as follows.

1. Perform a morphological analysis and remove noise.
2. Calculate the feature vector.
3. Classify by SVR.

C. Morphological analysis and noise removal

In messages posted on the BBS automatically collected by our program, sentences in Japanese are not divided into words as they are in English. Therefore, to extract words, we divide sentences into morphemes by using a morphological analyzer[5]. Next, to remove unsuitable words to obtain feature vector, we remove noise as follows.

1) Remove alphanumeric characters, symbols, and words other than Japanese ones.

2) Remove unnecessary words (particles, auxiliary verbs, conjunctions, attributive poetry, adverbs, numbers, and pronouns, interjections, and proper nouns).

3) Reflect negative words[6]

D. Extraction of useful words for classification and calculation of feature vector

To develop a dictionary, we manually extracted words thought to express sentiment among the top 5000 words in terms of the TF・IDF (term frequency, inverse document frequency) value calculated from messages about Sony posted on Yahoo! BBS in 2005, and we add this list of words to the semantic orientations of words extracted by Takamura et al. [10]. The number of words in our dictionary is 6989.

The feature vector of each message has 6989 dimensions, and each value is the importance of a word. We define the importance as the function of the appearance frequency of each word and calculate it by TF*IDF. TF increases the importance of a word that appears many times in a message. IDF is a filter of popular words because it increases the importance of a word that appears only in a specific message. The importance *w(t, d)* is given by

$$w(t,d) = \ln(tf_{t,d} + 1) \cdot \ln(\frac{N}{df_d})$$

$w(t,d)$: Importance

t : Message

d : Word

N : The number of messages posted in each period.

$tf_{t,d}$: An appearance frequency of word : d in message : t.

df_d : The number of messages including word : d.

Messages for which the feature vector cannot be extracted are excluded from our classification as noise. The feature vector of message f_t is defined as

$$f_t = (w(t,1), w(t,2), ..., w(t,l))$$

l : Total number of word (6989 words)

---

[5] We use Mecab (http://mecab.sourceforge.net/) as a morphological analyzer.

[6] A negative word denies the previous morpheme in Japanese. To reflect negative words, we add Japanese negative suffixes such as ナイ, ません, ず, and ぬ to the end of a word of the previous morpheme.

E. Classification by SVR

Using SVR, Okanohara and Tsujii [9] showed that recommendations can be estimated from reviews using the contents and recommendations of customer reviews on Amazon.co.jp (in Japan). We classify messages posted on the BBS using the sentiments and contents of messages.

Yahoo! BBS has a function for disclosing the poster's sentiment to the public. The poster of each message in Yahoo! BBS can select a sentiment from five alternatives: "strong buy", "buy", "hold", "sell", and "strong sell". However, a lot of messages are posted without sentiments. The total number of messages of the top 50 companies that we acquire is 1,441,251[7] during the analysis period. Of these the number for which we can extract the feature vector is 1,106,310. However, only 232,077 messages have the poster's sentiment (21%). Therefore, we classify all the messages without sentiments into "bullish", "neither", or "bearish" .

The learning data is messages with disclosed sentiments. Input data is the feature vectors of the messages, and the output for each message is "strong buy"=1, "buy"=0.5, "hold"=0, "sell"=-0.5, and "strong sell"=-1. Machine learning using SVR is done for all messages of all companies for which we can extract feature vectors. We classify messages having an SVR output equal to or larger than 0.5 as "bullish", equal or smaller than - 0.5 as "bearish", and other (from -0.5 to 0.5) as "neither". We use LibSVM[8] as an SVR program.

The classification results for the learning data during the analysis period are shown in Table 1. When the accuracy is taken to be the probability of "strong buy" and "buy" classified as "bullish", "hold" as "neither", and "strong sell" and "sell" as "bearish", the accuracy is 61.3%.

A breakdown shows that the probability of "strong buy" and "buy" being classified as "bullish" is 84.3%. The probability of "hold" being "neither" is 35.7%, but within this "bullish" is 57.8%. The probability of "strong sell" and "sell" being "bearish" is 37.9%, but that for "neither" was 41.7% and that for "bullish" is 20.4%. Therefore, there is a tendency for bullish to shift in our classification.

We can consider two reasons for the shift in bullishness. First, the contents of posted messages and their sentiments seemed to differ. Though details are omitted because of space limitations, there are messages whose sentiments seems not to correspond to the content. Therefore, we use the SVR classification results without any adjustment.

Second, there is a problem with the accuracy of our machine learning. From the viewpoint of accuracy, the probability of "strong buy" and "buy" messages being bullish is 68.2%. The probability of "strong sell" and "sell" messages being bearish is 78.1%. However, the probability of "hold" messages being neither is only 38.3%. Therefore, we use only data classified as bullish and bearish and do not use the data classified as neither.

[7] The number of messages that can actually be acquired for the top 50 companies (as of May 31, 2007).

[8] LibSVM is a library for SVM provided by Dr. Chih-Jen Lin. http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Using the above procedure, we classify 1,106,310 messages about 50 companies during the analysis period including messages without sentiments into bullish, neither, or bearish. The results are shown in Table 2. They show that 645,253 messages (58.3%) are bullish and 82,159 (7.4%) are bearish. This data is used for the following analyses.

TABLE I. SVR CLASSIFICATION RESULTS FOR 50 COMPANIES (DECEMBER 31, 2006 TO JANUARY 1, 2005).

		classification						total		accuracy
		bearish		neither		bullish				
sentiment	strong buy	591	(0.3%)	9,502	(4.1%)	82,282	(35.5%)	92,375	(39.8%)	84.3%
	buy	187	(0.1%)	8,671	(3.7%)	19,122	(8.2%)	27,980	(12.1%)	
	hold	4,181	(1.8%)	23,311	(10.0%)	37,721	(16.3%)	65,213	(28.1%)	35.7%
	sell	1,003	(0.4%)	4,932	(2.1%)	1,836	(0.8%)	7,771	(3.3%)	37.9%
	strong sell	16,633	(7.2%)	14,456	(6.2%)	7,649	(3.3%)	38,738	(16.7%)	
total		22,595	(9.7%)	60,872	(26.2%)	148,610	(64.0%)	232,077	(100.0%)	61.3%
accuracy		78.1%		38.3%		68.2%		61.3%		

accuracy is sum of

TABLE II. BULLISH-BEARISH CLASSIFICATION RESULTS.

		bearish (x<=-0.5)		neither (-0.5<x<0.5)		bullish (0.5<=x)		total
msg	Learning data	22,595	9.7%	60,872	26.2%	148,610	64.0%	232,077
	Without sentiment	59,564	6.8%	318,026	36.4%	496,643	56.8%	874,233
	total	82,159	7.4%	378,898	34.2%	645,253	58.3%	1,106,310

F. BBS indexes

The BBS indexes are the number of messages, number of bullish messages, number of bearish messages, bullishness, and agreement index. We calculate them for each time window except for holidays and weekends. For instance, the total number of messages (t) is equal to the number of before-market messages (t) plus the number of in-market messages (t) plus the number of after-market messages (t). The sample size of each time window is 23,789[9].

(1) Number of messages: the number of messages posted in a given time window.

(2) Number of bullish messages: the number of messages classified by SVR as bullish in each time window.

(3) Number of bearish messages: the number of messages classified by SVR as bearish in an each time window.

(4) Bullishness: the difference between the numbers of bullish and bearish messages in each time window. To correct the influence of a large number of messages, we use a natural logarithm. If the total number of bullish and bearish messages is less than 3, we assume that the bullishness is not clear, so we exclude these data from our analysis. As a result, the sample sizes are 7,572 for the before-market time window, 12,351 for the in-market window, and 14,818 for the after-market window.

$$\ln\left(\frac{1+\text{No. of bullish messages}}{1+\text{No. of bearish messages}}\right)$$

(5) Agreement index: the agreement index shows the level of agreement between bullish and bearish opinions in each time window. If the total number of bullish and bearish messages is less than 3, we assume that the agreement index is unclear and we exclude it from our analysis.

[9] The number of items of data about 50 companies that we can acquire from 2005 to 2006: 477 trading days ×50 =23,850.

$$\frac{|\text{No. of bullish messages} - \text{No. of bearish messages}|}{\text{No. of bullish messages} + \text{No. of bearish messages}}$$

VI. RESULT

We analyze the correlations between the stock market and the BBS indexes on trading days (shown in Tables 3, 4, and 5).

TABLE III. CORRELATION WITH STOCK RETURN

		return			excess return		
		yesterday	today	next day	yesterday	today	next day
No. of msgs	before mkt	-0.017 *		-0.015 *			-0.015 *
	in market			-0.014 *	0.016 *		-0.016 *
	after mkt		-0.024 **			-0.017 **	
No. of bullish msgs	before mkt						
	in market	0.019 **	0.020 **		0.025 **	0.025 **	
	after mkt						
No. of bearish msgs	before mkt	-0.040 **	-0.031 **		-0.040 **	-0.032 **	
	in market	-0.032 **	-0.068 **	-0.016 *	-0.034 **	-0.070 **	-0.021 **
	after mkt	-0.014 *	-0.058 **	-0.024 **	-0.014 *	-0.059 **	-0.025 **
bullish ness	before mkt	0.075 **	0.063 **		0.082 **	0.065 **	0.024 *
	in market	0.066 **	0.105 **	0.018 *	0.071 **	0.112 **	0.024 **
	after mkt	0.052 **	0.110 **	0.036 **	0.055 **	0.108 **	0.036 **
agreement index	before mkt	0.052 **	0.039 **		0.054 **	0.035 **	
	in market	0.042 **	0.090 **	0.019 *	0.040 **	0.089 **	0.026 **
	after mkt	0.050 **	0.095 **	0.035 **	0.049 **	0.087 **	0.033 **

TABLE IV. CORRELATION WITH VOLUME

		Volume		
		yesterday	today	next day
No. of msgs	before mkt	0.189 **	0.180 **	0.156 **
	in market	0.226 **	0.251 **	0.206 **
	after mkt	0.189 **	0.212 **	0.195 **
No. of bullish msgs	before mkt	0.189 **	0.182 **	0.158 **
	in market	0.224 **	0.248 **	0.208 **
	after mkt	0.192 **	0.216 **	0.201 **
No. of bearish msgs	before mkt	0.077 **	0.075 **	0.067 **
	in market	0.061 **	0.073 **	0.055 **
	after mkt	0.058 **	0.065 **	0.061 **
bullish ness	before mkt	0.108 **	0.101 **	0.083 **
	in market	0.149 **	0.172 **	0.144 **
	after mkt	0.129 **	0.147 **	0.137 **
agreement index	before mkt			
	in market	0.041 **	0.049 **	0.046 **
	after mkt	0.043 **	0.050 **	0.051 **

TABLE V. CORRELATION WITH VOLUME

		volatility			absolute return		
		yesterday	today	next day	yesterday	today	next day
No. of msgs	before mkt	0.180 **	0.171 **	0.153 **	0.157 **	0.125 **	0.096 **
	in market	0.257 **	0.254 **	0.237 **	0.205 **	0.234 **	0.142 **
	after mkt	0.217 **	0.218 **	0.209 **	0.151 **	0.206 **	0.148 **
No. of bullish msgs	before mkt	0.162 **	0.156 **	0.142 **	0.141 **	0.111 **	0.092 **
	in market	0.233 **	0.230 **	0.217 **	0.187 **	0.209 **	0.131 **
	after mkt	0.198 **	0.199 **	0.192 **	0.140 **	0.186 **	0.135 **
No. of bearish msgs	before mkt	0.089 **	0.085 **	0.075 **	0.081 **	0.068 **	0.045 **
	in market	0.116 **	0.117 **	0.108 **	0.088 **	0.115 **	0.067 **
	after mkt	0.092 **	0.094 **	0.091 **	0.055 **	0.099 **	0.067 **
bullish ness	before mkt	0.092 **	0.094 **	0.089 **	0.074 **	0.063 **	0.057 **
	in market	0.152 **	0.146 **	0.139 **	0.110 **	0.137 **	0.081 **
	after mkt	0.105 **	0.104 **	0.100 **	0.073 **	0.099 **	0.063 **
agreement index	before mkt						
	in market	0.031 **	0.032 **	0.033 **		0.021 *	
	after mkt	0.029 **	0.029 **	0.028 **	0.023 **		

* : 5% significant level, ** : 1% significant level.

VII. DISCUSSION

Here, we discuss whether the BBS indexes correspond to leading, coincident, or lagging indicators of the stock market for hypotheses 1–9. A leading indicator is defined as a statistically significant BBS index before the market opened with today's stock market, and all BBS indexes with next trading day's stock market. A coincident indicator is defined as a statistically significant BBS index at the same time window as the market. A lagging indicator is a statistically significant BBS index after the market window with today's market, and all BBS indexes with the previous day's stock market.

A. *H 1: The no. of messages has no correlation with the stock return.*

As a leading indicator, the number of before-market and in-market messages has a 5% significant correlation with the next day's return. The number of messages can estimate the fall in stock prices. This result does not support the hypothesis, so hypothesis 1 is rejected. As a lagging indicator, the number of after-market messages has a negative 1% significant correlation with today's return and the excess return. This result shows that when stock prices fall, posters may post messages in reaction to the stock prices.

The number of bullish messages actually has a positive 1% significant correlation with the stock return. Therefore, the number of bullish messages is a coincident and lagging indicator of the return. This shows that the number of bullish messages may reflect the rise in the previous day's and today's stock prices. The number of bearish messages and the return has a negative significant correlation except between before-market messages and the next day's return. In particular, as a leading factor, the number of bearish after-market messages has a negative 1% significant correlation with the next day's return. Bearish opinions after the market has closed may be related to the next day's fall in stock prices. In our analysis period, there is no bearish market but a lot of bullish opinions on the BBS. In such a situation, the small number of bearish opinions might have influenced the investment behavior of many private investors.

B. *H 2: The no. of messages correlates with trading volume.*

The number of messages, number of bullish messages, and number of bearish messages have a positive 1% significant correlation with the trading volume. This result supports hypothesis 2. They are leading, coincident, and lagging indicators of volume, and we conjecture that they show the level of investor interest in each stock. Therefore, the number of messages is a representative variable of the extent to which private investors using Internet are interested in the stock.

C. *H 3: The no. of messages correlates with volatility.*

The number of messages, number of bullish messages, and number of bearish messages have a positive significant correlation with volatility and the absolute value of stock price change. These results support hypothesis 3. They are leading, coincident, and lagging indicators of volatility and the absolute value of price change. In agreement with the results of [1], the number of messages may forecast volatility. This result suggests that messages posted on a BBS may influence the behavior of investors in the Japanese market.

D. *H 4: Bullishness is positively correlated with stock return.*

Bullishness had a significant correlation with stock return and excess return except between before-market bullishness and the next day's return. This result supports hypothesis 4. In particular, because after-market bullishness has a 1% significant correlation with the next day's return, the overnight

opinions of investors may influence the next day's investment behavior and stock price.

E. H 5: Bullishness is positively correlated with volume.

Bullishness has a positive 1% significant correlation with volume. Hypothesis 5 is supported. Bullishness is a leading, coincident, and lagging indicator of volume. From 2005 to 2006, the market is not bearish, but is a bullish or box market. When bullishness is high, investors who monitor the BBS pay attention to the stock. This might cause investment, and the trading volume might increase.

F. H 6: Bullishness is positively correlated with volatility.

Bullishness has a positive 1% significant correlation with volatility and with the absolute value of return. Hypothesis 6 is supported. Bullishness is a leading, coincident, and lagging indicator of volatility. As in section E above, it might cause investment, and volatility might increase.

G. H 7-9: The agreement index is not correlated with the stock return, and negative collerated wuth volume, and volatility.

The agreement index has a positive correlation with the stock return and excess return. These results do not support hypothesis 7, so hypothesis 7 is rejected.

The in-market and after-market agreement indexes has a positive 1% significant correlation with volume. However, [1] reported that the agreement index has a negative relation to volume. In a word, there is a tendency for volume to decrease when opinions are one-sided to either bullish/bearish in the US stock market. On the other hand, our results show that volume in Japan tended to increase when opinions are one-sided. This does not support hypothesis 8, so hypothesis 8 is rejected.

There is a 1% significant positive correlation between volatility and the in-market and after-market agreement indexes. According to [1], volatility increases when opinions are not one-sided but conflict. Thus, hypothesis 9 is rejected.

Volatility is the standard deviation of return in five days covering two days each side of today. The two data periods do not necessarily correspond. To get them to match, the correlation with absolute value of return is shown in Table 5 right column. Even in this case, the in-market agreement index and the absolute value of return has a 5% significant positive correlation. Thus, hypothesis 9 is rejected.

Hypotheses 7, 8, and 9 concerning the agreement index are rejected, contrary to the results of [1]. The stock market in Japan tends to rise from 2006 to 2007, which is our analysis period. Therefore, this might be a pseudo-correlation caused by a rising market, so we perform a regression analysis of bullishness and agreement index. The correlation is 0.579 and 1% significant. The regression analysis is as follows.

Agreement index＝0.135*bullishness + 0.625

For our data, many agreement indexes have values of 1; in other words, bullish not bearish data. As a result, because of the bullish market, the correlation between bullishness and the agreement index might be a pseudo-correlation. Therefore, we think that it is not possible to determine the effect of the agreement index. We think it is necessary to analyze data that covers a bearish market.

VIII. CONCLUSION

We analyze Yahoo! BBS in Japan for the top 50 companies mentioned in messages posted on Yahoo! BBS. The relations between the market return and the number of messages is not 1% significant. As a leading indicator, it might be difficult to predict the stock price return. However, when we analyze the contents of messages, we find that if there are many bearish messages posted after the market closed (1500–2400), the tendency of the next day's return is negative. The bullishness and agreement level of posted messages might reflect the return of stocks, and the number of messages and bullishness might be leading indicators of trading volume and volatility.

The effect of the agreement index can not be determined by our analysis. It requires a period that includes bearish markets. Further research is necessary on an analysis period that includes a bearish market. Moreover, an analysis of companies mentioned in a small number of messages is a research topic for the future. In addition, we uniformly analyze the messages of 50 companies by the same natural-language processing. However, the meanings of words posted about each company might differ. We believe that individual message analysis for each company is necessary in the future. In our analysis, we exclude "neither" because its classification accuracy is low. For better classification accuracy, Das and Chen [3] proposed a method of combining five natural-language processing techniques. Moreover, we do not analyze the context of word combinations. These are also future research topics.

REFERENCES

[1] W.Antweiler and Frank, M. Z., "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *J. Finance*, Vol. 59, No. 3, pp.1259-1294, 2004.

[2] N.Cristianini and Shawe-Talor, J., "*An Introduction to Support Vector Machines and other kernel-based learning methods*", Cambridge University Press, 2000.

[3] S.R. Das, Chen, M. Y., "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web", *Manage.Sci.*, Vol.53, No.9, pp. 1375–1388, 2007.

[4] M.Harris and Raviv, A., "Differences of Opinion Make a Horse Race," *Review of Financial Stud.*, Vol. 6, pp. 473-506, 1993.

[5] R.Tumarkin and Whitelaw, R.F., "News or Noise? Internet Postings and Stock Prices," *Financial Analysts J.*, Vol.57, pp.41-51, 2001.

[6] J. Van Bommel, "Rumors, " *J.Finance*, Vol. 58, No. 4, pp.1499-1519, 2003.

[7] P.D. Wyscocki, "Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards," Working paper, University of Michigan, Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=160170, 1999.

[8] ACCII, BLOG ranking 500 of firms listed on TSE, Dec., 2006. (in japanese)

[9] D.Okanohara, Tsujii, J., "Assigning Polarity Scores to Reviews Using Machine Learning Techniques" In Robert Dale, Kam-Fai Wong, Jian Su and Oi Yee Kwong (Eds.), Natural Language Processing - IJCNLP 2005. LNCS3651. Jeju Island, Korea, Springer-Verlag, October 2005.

[10] ,H Takamura, Inui, T.,Okumura, M., "Extracting Semantic Orientations Using Spin Model(Natural-Language Processing)", Trans. Inform. Process. Soc. of Japan, , Vol.47, No.2, pp. 627-637, 2006. (in japanese)

[11] K.Yamashita, Ishigami, T., Sato, T., "The relationship at the Internet forum between social concern and stock price change", *Proc. of Annual Conf. Japan Association for Social Informatics*, Vol.20 No.1, pp.237-240, 2005. (in japanese)